

# Towards Solving Imperfect Recall Games

Jiří Čermák and Branislav Bošanský

Department of Computer Science, Faculty of Electrical Engineering  
Czech Technical University in Prague  
{cermak, bosansky}@agents.fel.cvut.cz

## Abstract

Imperfect recall games represent dynamic interactions where players can forget previously known information, such as the exact history of played actions. Opposed to perfect recall games, where the players remember all information, imperfect recall games allow a concise representation of strategies. However, most of the existing algorithmic results are negative for imperfect recall games and many theoretical and computational results do not translate from perfect recall games. The goal of this paper is to (1) summarize the existing results regarding imperfect recall games and (2) extend these results to a restricted subclass of imperfect recall games termed *A-loss recall games*. Finally, (3) we emphasize the impact of these theoretical results on algorithms that compute (approximately) optimal strategies in extensive-form games and show that they cannot be easily extended to imperfect recall games.

## Introduction

Dynamic games with a finite number of moves are modeled as *extensive-form games* (EFGs). These games are visualized as game trees where nodes correspond to states of the game and edges to actions executed by players. This representation is general enough to model stochastic events and imperfect information when players are unable to distinguish among several states.

A natural assumption in extensive-form games is that each player perfectly remembers the history of played actions and all information gained during the course of the game. This assumption is known as *perfect recall* and it is required by most of the algorithms that compute optimal strategies in EFGs in order to guarantee correctness (or convergence). However, the main caveat of the perfect recall is that conditioning the strategies on the complete history increases the size of the strategy exponentially with the horizon. In order to simplify the strategy space, an abstraction removing certain information previously available to players can be used to reduce the size of the strategy. This approach has been used for computer poker for quite some time (Sandholm 2010). However, to sufficiently simplify the strategy space, such an abstracted game might need to violate the assumption of perfect recall and have *imperfect recall* (Waugh et al. 2009).

Solving imperfect recall games is a difficult problem that we address in this paper. We summarize several known results in imperfect recall games to provide a complete picture of this problem. First, we discuss the difference between mixed and behavioral strategies in imperfect recall games and argue that only behavioral strategies exploit the benefits of the imperfect recall. Second, we discuss computation of a best response since some classes of imperfect recall games may require randomized strategies in order to find the best response. Moreover, even if pure strategies are sufficient as a best response, finding such a pure best response can be more difficult (NP-hard) compared to the perfect recall case, where the problem is polynomial. The reason is that in general, it may not be sufficient to compute a best response by selecting the best action to play in each decision point. However, there exists a subclass of imperfect recall games, termed *A-loss recall games*, where each loss of memory of a player can be traced back to a loss of memory of the player's own actions. In A-loss recall games, finding a best response can be done using the same algorithm as in the perfect recall case. This property can be exploited when designing new more efficient algorithms for approximating maximin strategies (Bošanský et al. 2017) and thus the class of A-loss recall games presents an interesting subclass of imperfect recall games. Unfortunately, most of the computational difficulties hold also for A-loss recall games.

We describe a simple zero-sum example game due to Wichardt (2008) demonstrating that even a zero-sum A-loss recall game does not have to have a Nash equilibrium (NE) in behavioral strategies. Next, we show that many existing hardness results for imperfect recall games due to Koller and Meggido (1992) and Hansen (2007) also hold for A-loss recall games. On the positive side, we provide novel necessary and sufficient condition for the existence of a NE in behavioral strategies in A-loss recall games.

Finally, we analyze the state of the art approximate algorithm for solving perfect recall zero-sum EFGs, Counterfactual Regret Minimization (CFR; (Zinkevich et al. 2007)). While there exist subclasses of imperfect recall games, where CFR provably converges to an optimal strategy (Lanctot et al. 2012; Kroer and Sandholm 2016; Lisý, Davis, and Bowling 2016), this convergence cannot be generalized to A-loss recall games. We show that CFR can converge to an incorrect strategy even though a NE exists in

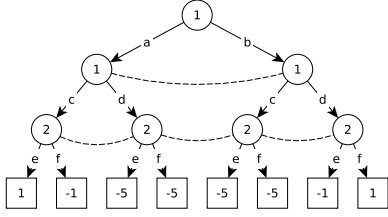


Figure 1: An EFG with imperfect recall, circle nodes represent the states of the game, numbers in the circles show which player acts in that node (player 1 or 2), dashed lines represent indistinguishable states and box nodes are the terminal states with utility values (player 1 maximizes the utility, player 2 minimizes). This game has A-loss recall, but a Nash equilibrium in behavioral strategies does not exist (from (Wichardt 2008)).

behavioral strategy. Moreover, this strategy can have an arbitrarily worse expected utility compared to the optimum.

### Extensive-Form Games

We first describe necessary technical introduction to extensive-form games (EFGs). A two player EFG  $G$  is a tuple  $\{\mathcal{P}, \mathcal{H}, \mathcal{Z}, P, u, \mathcal{I}, A\}$ .  $\mathcal{P} = \{1, 2\}$  denotes the set of players. We use  $i$  to refer to a player and  $-i$  to refer to the opponent of  $i$ . Set  $\mathcal{H}$  contains all the states of the game.  $P : \mathcal{H} \rightarrow \mathcal{P} \cup \{c\}$  is the function associating a player from  $\mathcal{P}$  or nature  $c$  with every  $h \in \mathcal{H}$ . Nature  $c$  represents the stochastic environment of the game.  $\mathcal{Z} \subseteq \mathcal{H}$  is a set of terminal states.  $u_i(z)$  is a utility function assigning to each leaf the value of preference for player  $i$ ;  $u_i : \mathcal{Z} \rightarrow \mathbb{R}$ . For zero-sum games it holds that  $u_i(z) = -u_{-i}(z), \forall z \in \mathcal{Z}$ . The imperfect information is defined using the information sets.  $\mathcal{I}_i$  is a partitioning of all  $\{h \in \mathcal{H} : P(h) = i\}$  into these information sets. All states  $h$  contained in one information set  $I_i \in \mathcal{I}_i$  are indistinguishable to player  $i$ . The set of available actions  $A(h)$  is the same  $\forall h \in I_i$ . We overload the notation and use  $A(I_i)$  as actions available in  $I_i$ . A game has *perfect recall* iff  $\forall i \in \mathcal{P} \forall I_i \in \mathcal{I}_i$  all the states in  $I_i$  share the same history for  $i$ . If there exists at least one information set where this does not hold, the game has *imperfect recall*.

### Strategies in Imperfect Recall Games

There are several representations of strategies in EFGs. A *pure strategy*  $s_i$  for player  $i$  is a mapping assigning  $\forall I_i \in \mathcal{I}_i$  a member of  $A(I_i)$ .  $\mathcal{S}_i$  is a set of all pure strategies for player  $i$ . A *mixed strategy*  $m_i$  is a probability distribution over  $\mathcal{S}_i$ , set of all mixed strategies of  $i$  is denoted as  $\mathcal{M}_i$ . *Behavioral strategy*  $b_i$  assigns a probability distribution over  $A(I_i)$  for each  $I_i$ .  $\mathcal{B}_i$  is a set of all behavioral strategies for  $i$ ,  $\mathcal{B}_i^p \subseteq \mathcal{B}_i$  denotes the set of deterministic behavioral strategies for  $i$ . A *sequence*  $\sigma_i$  is a list of actions of player  $i$  ordered by their occurrence on the path from the root of the game tree to some node. By  $seq_i(h)$  we denote the sequence of player  $i$  leading to the state  $h$ . A *strategy profile* is a set of strategies, one strategy for each player.

Alternatively, we can formulate a mixed strategy as a *realization plan*  $r_i$  that for a sequence  $\sigma_i$  represents the probability of playing actions in  $\sigma_i$  assuming the other players play such that the actions of  $\sigma_i$  can be executed. Realization plan  $r_i$  has to satisfy network flow property; i.e.,  $r_i(\sigma_i) = \sum_{a \in A(I_i)} r_i(\sigma_i \cdot a)$ , where  $I_i$  is an information set reached by sequence  $\sigma_i$  and  $\sigma_i \cdot a$  stands for  $\sigma_i$  extended by action  $a$ .  $\mathcal{R}_i$  is a set of realization plans for  $i$ ,  $\mathcal{R}_i^p \subseteq \mathcal{R}_i$  denotes the set of pure realization plans for  $i$ . A pair of strategies with arbitrary representation is *behaviorally equivalent* if they generate the same probability distribution over all  $z \in \mathcal{Z}$ . We overload the notation and use  $u_i$  as the expected utility of  $i$  when the players play according to pure (mixed, behavioral) strategies or realization plans.

Behavioral strategies, realization plans, and mixed strategies have all the same expressive power in perfect recall games, but it can differ in imperfect recall games (Kuhn 1953). Moreover, the size of these representations differs significantly. Mixed strategies of player  $i$  state probability distribution over  $\mathcal{S}_i$ , where  $|\mathcal{S}_i| \in \mathcal{O}(e^{|\mathcal{Z}|})$ , realization plans create a network flow over the set of sequences with size in  $\mathcal{O}(|\mathcal{Z}|)$ , and finally behavioral strategies create probability distribution over the set of actions (note that its size is proportional to the number of information sets, which can be exponentially smaller than  $|\mathcal{Z}|$ ). We will demonstrate the differences on the following example.

**Example:** Consider the game depicted in Figure 1. This game has 4 pure strategies for player 1  $\mathcal{S} = \{(a, c), (a, d), (b, c), (b, d)\}$ . A mixed strategy can condition the actions of players on information that the players should forget. For example, a mixed strategy where  $(a, c)$  and  $(b, d)$  are played with uniform probability 0.5 allows player 1 to condition playing  $c$  and  $d$  on the outcome of the non-deterministic choice in the root of the game, and thus randomize between the leftmost and the rightmost state in information set of player 2. Note that one cannot model the same behavior using a behavioral strategy that assigns for every decision point a probability distribution over the actions available, and therefore no additional information can be disclosed to the player.

Now we can define the Nash equilibrium in behavioral strategies.

**Definition 1.** We say that strategy profile  $b = \{b_i, b_{-i}\}$  is a Nash equilibrium in behavioral strategies iff  $\forall i \in \mathcal{P} \forall b_i^p \in \mathcal{B}_i^p : u_i(b_i, b_{-i}) \geq u_i(b_i^p, b_{-i})$ .

### Best Response Computation

One of the main computational components in algorithmic game theory is the problem of computing a best response. Formally, a strategy (e.g., pure, mixed, behavioral) is a best response if the expected utility is maximal compared to all other strategies from the particular class. To denote the best response regardless of the type of the strategy (i.e., regardless whether we consider mixed or behavioral strategies), we use the term *ex ante best response*.

In one-shot normal-form games (or matrix games), it is sufficient to consider a pure best response. However, this is

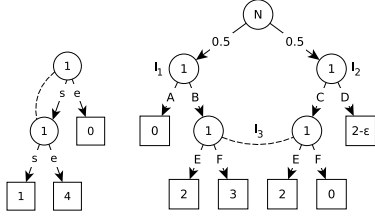


Figure 2: (Left) Absentminded Driver. (Right) An EFG without A-loss recall where a time consistent best response (playing the best action in an information set) is not necessarily the ex ante best response.

no longer true in imperfect recall games. Consider a one-player game called the absentminded driver (Piccione and Rubinstein 1997) depicted on the left in Figure 2. We see that by playing any pure, or even a mixed strategy, the player cannot reach outcome higher than 1. Note that in a mixed strategy a player samples from the set of all pure strategies; hence, the choice in each state of the information set in this game is the same. On the other hand, the ex ante optimal strategy is a behavioral strategy  $b(s) = \frac{4}{3}$  that reaches the expected value of  $\frac{4}{3}$ .

In general, we need to consider randomized strategies in so-called *absent minded* games, where there exists a path from the root of the game tree to some leaf such that an information set  $I_i \in \mathcal{I}_i$  is visited more than once. In the following text, we will assume that the games are without absentmindedness, where it is sufficient to consider only pure strategies to find an ex ante best response.

**Lemma 1.** *Let  $G$  be an imperfect recall game without absentmindedness and  $\beta_1$  strategy of player 1. There exists an ex ante pure behavioral best response of player 2.*

*Proof.* If the strategy of player 1 is fixed, then finding the best response for player 2 is finding an optimum of a function over a closed convex polytope with vertices formed by pure behavioral strategies. Since each action can be chosen at most once in a game without absentmindedness, the objective function is multilinear, therefore an optimum must be in one of the vertices of this polytope – that is a pure behavioral strategy.  $\square$

### Time Consistency and A-loss Recall

The problem of finding an ex ante best response in games without absentmindedness is still NP-hard (follows trivially from complexity results in (Koller and Megiddo 1992)) while the problem is easy (polynomial) in perfect recall games. The main difference is that in imperfect recall games an ex ante best response cannot be found by selecting an action with the highest expected utility to be played in each information set (called a *time consistent strategy* (Kline 2002)). This is caused by the fact that the belief in an information set of a player is not perfectly determined by the strategy of the opponent and nature, but also by the strategy of the best-responding player.

Consider the game in Figure 2 (right) between player 1 and chance. The ex ante best response of player 1 in this game is to play  $B, D, F$  getting the utility of  $\frac{5-\epsilon}{2}$ . Note, however, that when player 1 chooses the best action to play in each information set  $I$  separately given possible beliefs over states in  $I$ , one can find the strategy  $B, C, E$  with the expected utility of 2.

The equivalence between time consistent strategies and ex ante best responses is shown to hold in a subset of imperfect recall games called *A-loss recall games*. Consequently, the computation of the best response is in P in A-loss recall games (Kaneko and Kline 1995; Kline 2002).

**Definition 2.** *Player  $i$  has A-loss recall if and only if for every  $I \in \mathcal{I}_i$  and nodes  $h, h' \in I$  it holds either (1)  $seq_i(h) = seq_i(h')$ , or (2)  $\exists I' \in \mathcal{I}_i$  and two distinct actions  $a, a' \in \mathcal{A}_i(I')$ ,  $a \neq a'$  such that  $a \in seq_i(h) \wedge a' \in seq_i(h')$ .*

Condition (1) in the definition says that if player  $i$  has perfect recall then she also has A-loss recall. Condition (2) can be interpreted as requiring that each loss of memory of A-loss recall player can be traced back to some loss of memory of the player's own previous actions. Note that player 1 does not have A-loss recall in the game in Figure 2 (right) since parents of the nodes in information set  $I_3$  are in two distinct information sets  $I_1, I_2$  and their common predecessor is a chance node. On the other hand, the example game from Figure 1 has A-loss recall, which implies that the existence of NE is not guaranteed in A-loss recall games.

### Existence of NE in A-loss Recall Games

In the previous section, we have established that the existence of Nash equilibrium is not guaranteed, even in A-loss recall games. In this section, we provide both sufficient and necessary conditions for the existence of Nash equilibrium in behavioral strategies in two-player A-loss recall games.

First, we define a partition  $H(I_i)$  of states in every information set  $I_i$  of some imperfect recall game  $G$  to the largest possible subsets, not causing imperfect recall. More formally, let  $H(I_i) = \{H_1, \dots, H_n\}$  be a disjoint partition of all  $h \in I_i$ , where  $\bigcup_{j=1}^n H_j = I_i$  and  $\forall H_j \in H(I_i) \forall h_k, h_l \in H_j : seq_i(h_k) = seq_i(h_l)$ , additionally for all distinct  $H_k, H_l \in H(I_i) : seq_i(H_k) \neq seq_i(H_l)$ .

**Definition 3.** *The coarsest perfect recall refinement  $G'$  of the imperfect recall game  $G = \{\mathcal{P}, \mathcal{H}, \mathcal{Z}, P, u, \mathcal{I}, A\}$  is a tuple  $\{\mathcal{P}, \mathcal{H}, \mathcal{Z}, P, u, \mathcal{I}', A'\}$ , where  $\forall i \in \mathcal{P} \forall I_i \in \mathcal{I}_i H(I_i)$  defines the information set partition  $\mathcal{I}'$ .  $A'$  is a modification of  $A$ , which guarantees that  $\forall I \in \mathcal{I}' \forall h_k, h_l \in I A'(h_k) = A'(h_l)$ , while  $\forall I^k, I^l \in \mathcal{I}' A(I^k) \neq A(I^l)$ .*

In the Definition 3 we change the labelling of actions described by  $A$  to  $A'$ , since we modify the structure of the imperfect information  $\mathcal{I}$  to  $\mathcal{I}'$ . We denote by  $\Phi : E \rightarrow A'$  a function which for an edge  $e \in E$ , where  $E$  is the set of all edges of the game tree of  $G$  (and therefore of  $G'$ ), returns its action label in  $G'$ , similarly we define  $\Psi : E \rightarrow A$ . In the rest of this section, when we talk about the equivalence of arbitrary strategy representation or sequences in  $G$  and  $G'$ , we talk about the equivalence with respect to  $\Phi$  and  $\Psi$ . Same goes for applying the strategy from  $G$  to  $G'$  and vice versa.

**Lemma 2.** When player  $i$  plays according to a pure realization plan  $r_i$  in A-loss recall game  $G$  only states in one  $H_k \in H(I_i)$  in every imperfect recall information set  $I_i$  of  $i$  can have positive probability of occurrence if  $I_i$  gets visited.

*Proof.* This follows from the A-loss recall property.  $\square$

**Lemma 3.** The set of pure realization plans  $\mathcal{R}^{pG}$  of an arbitrary imperfect recall game  $G$  forms a subset of  $\mathcal{R}^{pG'}$  of its coarsest perfect recall refinement  $G'$ .

*Proof.* This holds since  $G$  is created from  $G'$  by merging information sets and therefore invalidating the mutually exclusive pure realization plans.  $\square$

**Lemma 4.** The set of pure realization plans  $\mathcal{R}^{pG}$  of an A-loss recall game  $G$  is equal to  $\mathcal{R}^{pG'}$  of its coarsest perfect recall refinement  $G'$ .

*Proof.* First, we prove that  $\forall i \in \mathcal{P} \forall r'_i \in \mathcal{R}^{pG'}, r'_i$  forms a pure realization plan of  $G$ . This follows from Lemma 2, as each  $H_k$  coincides with an information set in  $G'$  and  $i$  can choose actions independently in every  $H_k$  when following pure realization plan in  $G$ .  $r'_i$  must therefore be a valid pure realization plan of  $G$ .

Next, we show that  $\forall i \in \mathcal{P} \forall r_i \in \mathcal{R}^{pG}$  it holds that  $r_i$  forms a pure realization plan of  $G'$ . This directly follows from Lemma 3.  $\square$

**Lemma 5.** Every pure realization plan of an A-loss recall game  $G$  can be represented as a deterministic behavioral strategy and vice versa.

*Proof.* First, we show how to create a behavioral strategy  $b_i^p \in \mathcal{B}_i^{pG}$  equivalent to any  $r_i^p \in \mathcal{R}_i^{pG}$ . Moving from the root of the game tree of  $G$  for every  $I_i$ , for which  $r_i^p$  prescribes playing action  $a$ , we create  $b(I_i, a) = 1$ . This creates equivalent behavior in all  $I_i$ .

Creating pure realization plan from deterministic behavioral strategy is done by the same procedure, ignoring the parts of the tree unreachable when playing this strategy.  $\square$

Now we are ready to state the main result of this section.

**Theorem 1.** An A-loss recall game  $G$  has a Nash equilibrium in behavioral strategies if and only if there exists a Nash equilibrium  $b$  in behavioral strategies of the coarsest perfect recall refinement  $G'$  of  $G$ , such that  $\forall I \in \mathcal{I}^G \forall H_k, H_l \in H(I) : b(H_k) = b(H_l)$ , where  $b(H)$  stands for the behavioral strategy in the information set of  $G'$  formed by states in  $H$ .

*Proof.* First, since  $b$  is a Nash Equilibrium of  $G'$  we know that there exists no incentive for any player to deviate to any pure behavioral strategy in  $G'$ . From the interchangeability of pure behavioral strategies and pure realization plans in games with perfect recall (Kuhn 1953), we know that there exists no pure realization plan to which any of the players wants to deviate. From Lemma 3, it follows that there can exist no pure realization plan in  $G$ , and therefore by Lemma

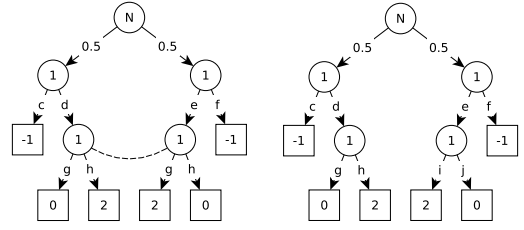


Figure 3: (Left) An imperfect recall game without A-loss recall where the only Nash equilibrium in behavioral strategies is not a Nash equilibrium of the coarsest perfect recall refinement. (Right) Its coarsest perfect recall refinement.

5 no pure behavioral strategy in  $G$  to which any of the players want to deviate. This, in combination with the fact, that  $b$  prescribes valid strategy in  $G'$  implies that  $b$  is a Nash Equilibrium in behavioral strategies of  $G$ .

Second, we prove that there exists no Nash equilibrium  $b'$  in behavioral strategies of  $G$  which is not a Nash equilibrium of  $G'$ . Let us assume that such  $b'$  exists. This would imply that there is no pure behavioral strategy in  $G$  to which players want to deviate when playing according to  $b'$ , and therefore no pure realization plan (Lemma 5). However, then Lemma 4 implies that there is no pure realization plan and therefore pure behavioral strategy in  $G'$  either, implying that  $b'$  is a Nash Equilibrium in  $G'$ . This contradicts the assumption and completes the proof.  $\square$

Informally, Theorem 1 states that  $G$  has a Nash equilibrium in behavioral strategies if and only if there exists a behavioral Nash equilibrium  $b$  in  $G'$  which prescribes the same behavior in every information set which is connected to some imperfect recall information set of  $G$ .

**Corollary 1.** An imperfect recall game  $G$  (not restricted to A-loss recall) has a Nash equilibrium in behavioral strategies if there exists a Nash equilibrium  $b$  in behavioral strategies of the coarsest perfect recall refinement  $G'$  of  $G$ , such that  $\forall I \in \mathcal{I}^G \forall H_k, H_l \in H(I) : b(H_k) = b(H_l)$ , where  $b(H)$  stands for the behavioral strategy in the information set of  $G'$  formed by states in  $H$ , the opposite does not hold.

*Proof.* The proof follows directly from the first part of the proof of Theorem 1, since every pure behavioral strategy in  $G$  can be expressed as a pure realization plan.

Finally, we provide a counter-example showing that the opposite direction of the implication does not hold. Consider the game in left subfigure of Figure 3. Here the only Nash equilibrium in behavioral strategies is playing  $d$  and  $e$  deterministically and mixing uniformly between  $g$ ,  $h$ . The only Nash equilibrium of its coarsest perfect recall refinement (shown in right subfigure of Figure 3) is, however, playing  $d$ ,  $e$ ,  $h$  and  $i$  deterministically.  $\square$

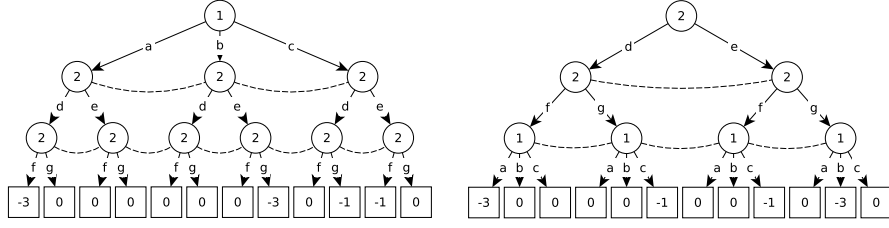


Figure 4: (Left) An imperfect recall game where maxmin strategy requires irrational numbers (Koller and Megiddo 1992). (Right) An A-loss recall game where maxmin strategy requires irrational numbers.

## Representation of Nash and Maxmin Strategies

In this section, we state first negative results concerning strategy representation in imperfect recall games and show that this issue is present also in A-loss recall games. In two-player zero-sum perfect recall games with rational payoffs, there exists a maxmin behavioral strategy which uses only rational probabilities (Koller and Megiddo 1992). In imperfect recall games, this no longer holds (Koller and Megiddo 1992). We present the example of the game where maxmin strategy requires irrational numbers. We transform this example to A-loss recall game where the maxmin strategy still needs irrational numbers, moreover, since the maxmin strategy is a part of Nash equilibrium of this game, we extend this result also to NE of A-loss recall games. It follows that computing exact maxmin and NE strategies is fundamentally difficult even in A-loss recall games.

**Theorem 2.** *The maxmin strategy may require irrational numbers, even in two-player zero-sum imperfect recall game with rational payoffs (Koller and Megiddo 1992).*

*Proof.* The example of the imperfect recall game used in (Koller and Megiddo 1992) is depicted in Figure 4 (left). The maxmin strategy of player 2 is trying to maximize

$$\min\{3b_2(d)b_2(f), 3(1 - b_2(d))(1 - b_2(f)), b_2(d)(1 - b_2(f)) + (1 - b_2(d))b_2(f)\}. \quad (1)$$

This is maximized when

$$3b_2(d)b_2(f) = 3(1 - b_2(d))(1 - b_2(f)) \\ = b_2(d)(1 - b_2(f)) + (1 - b_2(d))b_2(f), \quad (2)$$

which leads to  $b(d) = 0.1(5 \pm \sqrt{5})$ .  $\square$

**Theorem 3.** *The maxmin strategy may require irrational numbers, even in two-player zero-sum A-loss recall game with rational payoffs.*

*Proof.* In Figure 4 (right) we present an A-loss recall game obtained from the game in Figure 4 (left) by simply switching the order of players, which keeps the maxmin behavior explained above.  $\square$

**Theorem 4.** *The Nash equilibrium strategies may require irrational numbers, even in two-player zero-sum A-loss recall game with rational payoffs.*

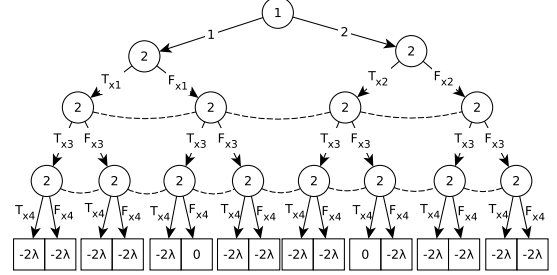


Figure 5: An imperfect recall game reduction from Theorem 5 of 3-SAT problem  $x_1 \vee \neg x_3 \vee x_4 \wedge \neg x_2 \vee x_3 \vee \neg x_4$ .

*Proof.* Strategy profiles  $b(d) = 0.1(5 \pm \sqrt{5})$ ,  $b(f) = 1 - 0.1(5 \pm \sqrt{5})$ ,  $b(a) = b(b) = 0.2$  and  $b(d) = 1 - 0.1(5 \pm \sqrt{5})$ ,  $b(f) = 0.1(5 \pm \sqrt{5})$ ,  $b(a) = b(b) = 0.2$  form all the Nash equilibria of the game in Figure 4 (right).  $\square$

## Computational Complexity in Imperfect Recall Games

Now we turn to the computational complexity of solving imperfect recall games. We present known results of Koller and Megiddo (1992) showing that computing maxmin strategies is NP-hard in imperfect recall games and that it is NP-hard to decide whether there exists a Nash equilibrium of behavioral strategies in imperfect recall games (Hansen, Miltersen, and Sørensen 2007). We further show, that both negative results directly translate to A-loss recall games.

**Theorem 5.** *The problem of deciding whether player 2 having an imperfect recall can guarantee an expected payoff of at least  $\lambda$  is NP-hard even if player 1 has perfect recall, there are no chance moves and the game is zero-sum (Koller and Megiddo 1992).*

*Proof.* The proof is made by reduction from 3-SAT problem. Given  $n$  clauses  $x_{j,1} \vee x_{j,2} \vee x_{j,3}$  we create a two person zero-sum game in a following way. In the root of the game player 1 chooses between  $n$  actions, each corresponding to one clause. Each action of player 1 leads to a state of player 2. Every such state of player 2 corresponds to the variable  $x_{j,1}$  where  $j$  is the index of the clause chosen in the root

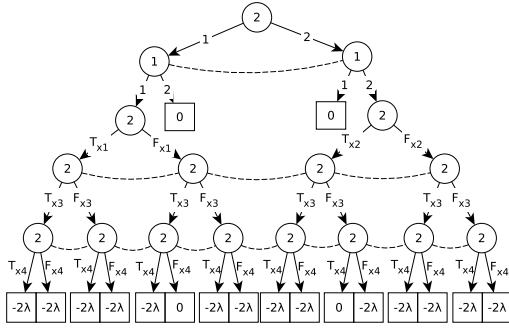


Figure 6: An A-loss recall game reduction from Theorem 6 of 3-SAT problem  $x_1 \vee \neg x_3 \vee x_4 \wedge \neg x_2 \vee x_3 \vee \neg x_4$ .

of the game. Every such state has actions  $T_{x_{j,1}}, F_{x_{j,1}}$  available, these actions correspond to setting the variable  $x_{j,1}$  to true or false respectively. After both  $T_{x_{j,1}}, F_{x_{j,1}}$  in  $x_{j,1}$  we reach state representing the assignment to  $x_{j,2}$  with the same setup (state representing the assignment to  $x_{j,3}$  is reached after that). After the assignment to  $x_{j,3}$  we reach the terminal state with utility  $(-\lambda, \lambda)$  if the assignment to  $x_{j,1}, x_{j,2}$  and  $x_{j,3}$  satisfies the clause  $x_{j,1} \vee x_{j,2} \vee x_{j,3}$ ,  $(0, 0)$  otherwise. The information sets of player 2 group together all the states corresponding to the assignment to one variable in the original 3-SAT problem (note that we assume that the order of variables in every clause follows some complete ordering on the whole set of variables in the 3-SAT problem). An example of the reduction is shown in Figure 5.

When the reduction is done in this way, we are sure that player 2 can guarantee the outcome of  $\lambda$  if and only if there is an assignment of variables satisfying the original 3-SAT problem. This holds since if there would be at least one clause not fully satisfied, player 1 could choose an action corresponding to this clause, guaranteeing an expected outcome smaller than  $\lambda$  for player 2. On the other hand, if player 2 can guarantee an outcome of  $\lambda$  all the clauses must be satisfiable since otherwise player 1 could again choose the action corresponding to some not fully satisfied clause given current strategy of player 2. The reduction is polynomial, since the game has  $2^3 n$  leafs.  $\square$

**Theorem 6.** *The problem of deciding whether player 2 having an A-loss recall can guarantee an expected payoff of at least  $\lambda$  is NP-hard even if player 1 has perfect recall, there are no chance moves and the game is zero-sum.*

*Proof.* The proof is made by reduction from 3-SAT problem. We modify the original reduction of Koller shown above in the following way. In the root of the game player 2 chooses between  $n$  actions, each corresponding to one clause. Player 1 plays next with no information about the action chosen by player 2. He has again  $n$  actions available. In every state of player 1  $n - 1$  actions lead directly to a terminal state with utility  $(0, 0)$  and one action (corresponding to the same clause as the action of player 2 preceding this

state) leads to a state of player 2. Every such state of player 2 corresponds to the variable  $x_{j,1}$  where  $j$  is the index of the clause chosen in the root of the game. The rest of the game is identical to the previous reduction with exception of utility values – every terminal state has assigned utility  $(-n\lambda, n\lambda)$  if the corresponding assignment to  $x_{j,1}, x_{j,2}$  and  $x_{j,3}$  satisfies the clause  $x_{j,1} \vee x_{j,2} \vee x_{j,3}$ ,  $(0, 0)$  otherwise.

When the reduction is done in this way, we are again sure that player 2 can guarantee the outcome of  $\lambda$  if and only if there is an assignment of variables satisfying the original 3-SAT problem. This holds since if there would be at least one clause not fully satisfied, player 1 could choose an action corresponding to this clause, guaranteeing an expected outcome smaller than  $\lambda$  for player 2. On the other hand, if player 2 can guarantee an outcome of  $\lambda$  all the clauses must be satisfiable since otherwise player 1 could again choose the action corresponding to some not fully satisfied clause given current strategy of player 2. The reduction is polynomial, since the game has  $n(n - 1) + 2^3 n$  leafs.

The last thing which remains to be shown is that player 2 has A-loss recall. This is satisfied since any loss of information about actions of player 1 can be tracked back to forgetting his own action taken in the root.  $\square$

**Theorem 7.** *It is NP-hard to check whether there exists a Nash equilibrium in behavioral strategies in two player imperfect recall games (Hansen, Miltersen, and Sørensen 2007).*

*Proof.* The proof is by reduction from 3-SAT. Given a 3-CNF formula  $F$  with  $n$  clauses we construct a zero-sum two-player game  $G$  as follows. Player 1 (the max-player) starts the game by making two actions, each time choosing a clause of  $F$ . We put all corresponding  $n + 1$  nodes (the root plus  $n$  nodes in the next layer) in one information set. If he fails to choose the same clause twice, he receives a payoff of  $-n^3$  and the game stops. Otherwise, the game continues in the same way as in the proof of Theorem 5. If the choices of player 2 satisfy the clause, player 1 receives payoff 0. If none of them do, Player 1 receives payoff 1.

The proof is now concluded by the following claim:  $G$  has an equilibrium in behavior strategies if and only if  $F$  is satisfiable. Assume first that  $F$  is satisfiable.  $G$  then has the following equilibrium (which happens to be pure): Player 2 plays according to a satisfying assignment while player 1 uses an arbitrary pure strategy. The payoff is 0 for both players and no player can modify their behavior to improve this so we have an equilibrium. Next, assume that  $G$  has an equilibrium. We shall argue that  $F$  has a satisfying assignment. We first observe that player 1 in equilibrium must have expected payoff at least 0. If not, he could switch to an arbitrary pure strategy and would be guaranteed a payoff of at least 0. Now look at the two actions (i.e., clauses) that player 1 is most likely to choose. Let clause  $i$  be the most likely and let clause  $j$  be the second-most likely. If player 1 chooses  $i$  and then  $j$  he gets a payoff of  $-n^3$ . His maximum possible payoff is 1 and his expected payoff is at least 0. Hence, we must have that  $-n^3 p_i p_j + 1 \geq 0$ . Since  $p_i \geq \frac{1}{n}$ , we have that  $p_j \leq \frac{1}{n^2}$ . Since clause  $j$  was the second most likely choice,

we in fact have that  $p_i \geq 1 - (n-1)(\frac{1}{n^2}) > 1 - \frac{1}{n}$ . Thus, there is one clause that player 1 plays with probability above  $1 - \frac{1}{n}$ . Player 2 could then guarantee an expected payoff of less than  $\frac{1}{n}$  for Player 1 by playing any assignment satisfying this clause. Since we are actually playing an equilibrium, this would not decrease the payoff of player 1 so player 1 currently has an expected payoff less than  $\frac{1}{n}$ . Now look at the assignment defined by the most likely choices of player 2 (i.e., the choices he makes with probability at least 0.5, breaking ties in an arbitrary way). We claim that this assignment satisfies  $F$ . Suppose not. Then there is some clause not satisfied by  $F$ . If Player 1 changes his current strategy to the pure strategy choosing this clause, he obtains an expected payoff of at least  $(\frac{1}{2})^3 \geq \frac{1}{n}$  (supposing, wlog, that  $n \geq 8$ ). This contradicts the equilibrium property and we conclude that the assignment, in fact, does satisfy  $F$ .  $\square$

**Theorem 8.** *It is NP-hard to check whether there exists a Nash equilibrium in behavioral strategies in two-player A-loss recall games.*

*Proof.* The proof is made by reduction from the 3-SAT problem. The reduction is similar to the one in proof of Theorem 6. The only change in the resulting game is the substitution of the utility in the leafs after actions of player 1 by  $(-1, 1)$  and in the leafs corresponding to satisfying the given clause by  $(-0.5, 0.5)$ .

When the 3-SAT problem is satisfiable, this game has a Nash equilibrium where both players mix uniformly in first two levels of the game, and the player 2 plays according to the assignment of variables satisfying this problem. If the 3-SAT problem is unsatisfiable, however, the NE does not exist. To show this, let us first denote as  $H_x$  the nodes of player 2 reached directly by actions of player 1. By  $u_2^h(b_1, b_2)$  we refer to the expected value of player 2 when starting in  $h \in H_x$  and playing according to behavioral strategies  $b_1$  and  $b_2$  (notice that  $u^h(b_1, b_2) \leq 0.5$  for any  $b_1, b_2$ ). To see that there is no NE, consider the situation where player 2 plays such strategy  $b_2$  (using subset of actions  $A'_2 \subseteq A_2(\text{root})$  in the root of the game) that player 1 is indifferent between actions from set  $A'_1 \subseteq A_1$ . In order to make sure this is a NE player 1 must play such  $b_1$  that player 2 does not want to deviate from  $b_2$ , while using only  $a_1 \in A'_1$  (otherwise  $b_1$  would not be stable). Note that the actions in  $A'_1$  and  $A'_2$  must correspond to the same clauses since player 1 will always prefer to play actions which lead to some  $h \in H_x$  with a positive probability.

First, let us assume that  $A'_2 \subset A_2(\text{root})$ . When playing such  $b_1$  and  $b_2$ , there are always members of  $H_x$  visited with positive probability. This implies that player 2 is better off by switching to  $b'_2$  which plays some  $a_2 \in A_2(\text{root}) \setminus A'_2$  deterministically, as  $u_2(b_1, b'_2) = 1 > u_2(b_1, b_2)$ .

Finally, let us assume that  $A'_2 = A_2(\text{root})$ . Since the original 3-SAT problem is not satisfiable it must hold that  $\forall b_1 \in \mathcal{B}_1, \forall b_2 \in \mathcal{B}_2 \exists h \in H_x u_2^h(b_1, b_2) < 0.5$ . Since we now force  $b_1$  and  $b_2$  to play all  $a_1 \in A_1$  and all  $a_2 \in A_2(\text{root})$  with positive probability (player 1 has to use actions corresponding to the same clauses as actions in  $A'_2$  to be able to make  $b_2$  stable), we know that such

$h$  is visited with positive probability. For all  $b_1$  playing all  $a_1 \in A_1$  with positive probability, therefore, always exists  $b'_2$  which plays some  $a_2 \in A'_2$  deterministically and which makes the clause corresponding to  $a_2$  satisfiable, for which  $u_2(b_1, b'_2) > u_2(b_1, b_2)$ . As shown above, however, there exists no pure strategy Nash equilibrium either.  $\square$

## CFR in Imperfect Recall Games

In this Section we discuss the problems which prevent the convergence of CFR in imperfect recall games. First, we briefly describe the ideas behind external regret and the Counterfactual regret minimization algorithm (Zinkevich et al. 2007). Finally, we discuss why CFR does not have to converge to no-regret strategies in imperfect recall games and show an example where CFR can converge to a strategy profile with arbitrarily worse expected value than the maxmin value of the game.

### External Regret

Given a sequence of behavioral strategy profiles  $b^1, \dots, b^T$ , the external regret for player  $i$ ,

$$R_i^T = \max_{b'_i \in \mathcal{B}_i} \sum_{t=1}^T (u_i(b'_i, b_{-i}^t) - u_i(b_i^t, b_{-i}^t)), \quad (3)$$

is the amount of utility player  $i$  could have gained if he played the best possible strategy across all time steps  $t \in \{1, \dots, T\}$ . An algorithm is a no-regret algorithm for player  $i$ , if the average positive regret approaches zero; i.e.  $\lim_{T \rightarrow \infty} R_i^{T,+}/T = 0$ , where  $R_i^{T,+} = \max(R_i^T, 0)$ .

### Counterfactual Regret Minimization

Counterfactual regret is defined in each iteration  $t$ ,  $I \in \mathcal{I}$  and  $a \in A(I)$  as  $r_i^t(I, a) = v_i(b_{I \rightarrow a}^t, I) - v_i(b^t, I)$ , where  $b_{I \rightarrow a}^t$  is the profile  $b^t$  except for  $I$ , where  $a$  is played and

$$v(b, I) = \sum_{z \in \mathcal{Z}_I} u_i(z) \pi_{-i}^b(z[I]) \pi^b(z[I], z), \quad (4)$$

where  $\pi^b(h)$  is the probability that  $h$  will be reached when players play according to the strategy profile  $b$ , with  $\pi_i^b$  and  $\pi_{-i}^b$  being the contribution of player  $i$  and all the players except  $i$ .  $z[I]$  is state  $h$  which is visited in  $I$  in order to reach leaf  $z$ . Finally,  $\pi^b(h, h')$  stands for the probability that  $h'$  will be reached from  $h$  when players play according to  $b$ . The immediate counterfactual regret is defined as  $R_{i, \text{imm}}^T(I) = \frac{1}{T} \max_{a \in A(I)} \sum_{t=1}^T r_i^t(I, a)$ . In games having perfect recall, minimizing the immediate counterfactual regrets at every information set minimizes the average external regret. This holds because perfect recall implies that the regret is bounded by the sum of the positive parts of the immediate counterfactual regrets (Zinkevich et al. 2007),

$$R_i^T \leq \sum_{I \in \mathcal{I}_i} \max_{a \in A(I)} R_{i, \text{imm}}^T(I, a), \quad (5)$$

and thus

$$\frac{R_i^T}{T} \leq \frac{\Delta_i |\mathcal{I}_i| \sqrt{|\mathcal{A}_i|}}{T}, \quad (6)$$

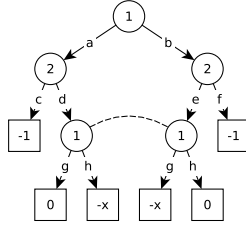


Figure 7: An A-loss recall game where CFR finds a strategy with the expected utility arbitrarily worse than the maximin value.

where  $|A_i| = \max_{I \in \mathcal{I}_i} |A(I)|$ . In imperfect recall games, however, equations (5) and (6) are not guaranteed to hold.

The no-regret learning cannot work in general in imperfect recall games, since the loss function  $l^t(b_i) = u_i(b_i^t, b_{-i}^t) - u_i(b_i, b_{-i}^t)$  used in equation (3) can be non-convex over the probability simplex of behavioral strategies (the loss function must be convex for a no-regret learning to have convergence guarantees (Gordon 2006)).

Consider the game from Figure 7. Let's assume we are in the step  $T$  of a no-regret learning algorithm and we evaluate the loss of some strategy  $b_1$  in step  $t < T$ . Now let us assume that  $b_1^t(a) = b_1^t(g) = 0.5$  and  $b_2^t(d) = b_2^t(e) = 1$ . Let  $b_1(a) = b_1(g) = 1$ ,  $b_1(b) = b_1(h) = 1$ , and finally  $b_1''(a) = b_1''(g) = 0.5$ . The losses of these strategies are  $l^t(b_1) = -x$ ,  $l^t(b_1') = -x$ ,  $l^t(b_1'') = 0$ . Since  $b_1''$  was created as a convex combination of  $b_1$  and  $b_1'$  with uniform weights, we immediately see that the loss function is non-convex, and therefore the convergence guarantees used in CFR due to Gordon (2006) no longer apply.

Notice that this is not the case in perfect recall games, since the behavior of  $i$  after any  $a, a' \in A(I_i)$  is independent  $\forall I \in \mathcal{I}_i$ .

Consider again the game in Figure 7. By increasing  $x > 2$  we can force the CFR to converge to a strategy with expected value arbitrarily worse than the maximin value -1. This holds, since mixing between actions  $a$  and  $b$  can yield the expected value strictly worse than the expected value reached by deterministic samples containing these actions if player 2 plays  $d$  and  $e$  with positive probability (the issue of non-convexity described above). It can, therefore, produce an average strategy, which can mix between actions evaluated when played deterministically only. Since the loss function can be non-convex for the mix, however, the average strategy has no guarantee for its expected value. We have experimentally evaluated our claim using outcome sampling CFR and we indeed observed convergence to mixed behavioral strategy. Vanilla CFR, on the other hand, does not specify how to compute the belief in the imperfect recall information set of  $i$ . Incorporating the strategy of player  $i$  to the belief does not guarantee convergence, since e.g., when player 1 plays a uniform strategy and player 2 plays  $d, e$ , vanilla CFR will perform no update, since all  $r_1^t(I, a)$  are zero and player 2 has negative regret for  $c$  and  $f$ .

This is an A-loss recall game with 2 Nash equilibria,

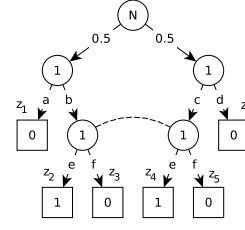


Figure 8: A well-formed game which does not have A-loss recall.

namely playing  $(a, g)$  or  $(b, h)$  deterministically for player 1 and  $(c, f)$  for player 2 (no mix between these two strategies for player 1 is a Nash equilibrium).

### (Skew) Well-Formed Games

A class of games where the CFR is still guaranteed to converge to  $(\epsilon)$ -optimal strategies is called ((chance relaxed) skew) well-formed games (Lanctot et al. 2012; Kroer and Sandholm 2016). This class of games restricts the structure of imperfect recall by requiring similarity of the states included in one imperfect recall information set both in the structure of past and future moves as well as the utilities reachable from the states. The result of such restriction is the bounding of the non-convexity of the loss function, such that the  $(\epsilon)$ -optimality is still attained.

**Definition 4.** We say that the imperfect recall game  $G$  is well-formed with respect to some perfect recall refinement  $G'$  if for all  $i \in \mathcal{P}$ ,  $I \in \mathcal{I}_i$ ,  $I', I''$  (where  $I', I''$  are information sets in  $G'$  which are unified to  $I$ ) there exists a bijection  $\alpha : \mathcal{Z}_{I'} \rightarrow \mathcal{Z}_{I''}$  and constants  $k_{I', I''}, l_{I', I''} \in [0, \infty)$  such that for all  $z \in \mathcal{Z}_{I'}$ :

1.  $u_i(z) = k_{I', I''} u_i(\alpha(z))$ ,
2.  $\pi_c(z) = l_{I', I''} \pi_c(\alpha(z))$ ,
3. in  $G$ ,  $X_{-i}(z) = X_{-i}(\alpha(z))$ , and
4. in  $G$ ,  $X_i(z[I'], z) = X_i(\alpha(z)[I''], \alpha(z))$ ,

where  $\mathcal{Z}_I$  stands for terminal states reachable from states in  $I$  and  $X_i(h)$  stands for the sequence of pairs  $(I, a)$ ,  $I \in \mathcal{I}_i$ ,  $a \in A(I)$  that occur when moving from the root of the game to  $h$ . We say that  $G$  is well-formed game if it is well-formed with respect to some perfect recall refinement.

Compared to A-loss recall games the (skew) well-formed games are significantly more restrictive since A-loss recall games restrict the structure of the game only above the imperfect recall information set (the requirement of forgetting only information of player's own actions) while the well-formed games restrict the structure above, below and also the structure of the utilities.

**Lemma 6.** In games with no chance, the well-formed games form a subset of A-loss recall games.

*Proof.* The only requirement of A-loss recall games is that players forget only their own actions, hence the restriction in information set  $I$  always concerns only the part of the



tree above  $I$ . We, therefore, focus on condition 3 in the definition of well-formed games, since it is the only one restricting the upper part of the game tree. Condition 3 requires that for each  $h' \in I'$  there must exist  $h'' \in I''$  such that  $X_{-i}(h') = X_{-i}(h'')$ , which, combined with the fact that there is no chance player, implies that there must exist difference in  $X_i(h')$  and  $X_i(h'')$ . Furthermore, since  $X_{-i}(h') = X_{-i}(h'')$  we are sure that there must exist  $(\bar{I}, \bar{a}) \in X_i(h')$  and  $(\hat{I}, \hat{a}) \in X_i(h'')$  such that  $\bar{I} = \hat{I}$  but  $\bar{a} \neq \hat{a}$ , which is exactly the condition 2 in the A-loss recall property. The rest of the requirements of the well-formed games considers utilities and parts of the tree not restricted in A-loss recall games, therefore the well-formed games form a subset of A-loss recall games.  $\square$

Notice that Lemma 6 holds also for skew well-formed games and chance relaxed skew well-formed games, since they provide relaxations in conditions (1) and (2) only.

Finally, in Figure 8 we show a well-formed game (with  $\alpha : \{z_2 \rightarrow z_4, z_3 \rightarrow z_5\}$ ) with chance which does not have A-loss recall.

## Conclusion

We summarize the current state of the art in the problem of solving imperfect recall games. We showed differences in the descriptive power of different strategy representations and difficulties in the best response computation. We described a subclass of imperfect recall games called A-loss recall games, where computation of best responses is significantly easier. However, we show that the negative results, such as no guarantee of the existence of Nash equilibria, NP-hardness of computation of Nash equilibrium and maxmin strategies and numerical issues in the representation of optimal strategies translate from imperfect recall to A-loss recall games. On the positive side, we show that the specific structure of A-loss recall games allows us to state sufficient and necessary conditions for the existence of Nash equilibrium in behavioral strategies. Finally, we show that applying no-regret algorithms with an aim to approximate optimal behavior has fundamental issues even in two-player zero-sum A-loss recall games where a Nash equilibrium exists.

## Acknowledgments

This research was supported by the Czech Science Foundation (grant no. 15-23235S), and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/235/OHK3/3T/13. We would like to thank the anonymous reviewers from WINE conference for their useful comments and counterexamples, some of which appear in this paper.

## References

Bosansky, B.; Cermak, J.; Horak, K.; and Pechoucek, M. 2017. Computing Maxmin Strategies in Extensive-Form Zero-Sum Games with Imperfect Recall. In *ICAART*.

Gordon, G. J. 2006. No-regret algorithms for online convex programs. In *NIPS*, 489–496.

Hansen, K. A.; Miltersen, P. B.; and Sørensen, T. B. 2007. Finding Equilibria in Games of no Chance. In *Computing and Combinatorics*. Springer. 274–284.

Kaneko, M., and Kline, J. J. 1995. Behavior Strategies, Mixed Strategies and Perfect Recall. *International Journal of Game Theory* 24:127–145.

Kline, J. J. 2002. Minimum Memory for Equivalence between Ex Ante Optimality and Time-Consistency. *Games and Economic Behavior* 38:278–305.

Koller, D., and Megiddo, N. 1992. The Complexity of Two-person Zero-sum Games in Extensive Form. *Games and Economic Behavior* 4:528–552.

Kroer, C., and Sandholm, T. 2016. Imperfect-Recall Abstractions with Bounds in Games. In *EC*, 459–476. ACM.

Kuhn, H. W. 1953. Extensive Games and the Problem of Information. *Annals of Mathematics Studies*.

Lancot, M.; Gibson, R.; Burch, N.; Zinkevich, M.; and Bowling, M. 2012. No-regret Learning in Extensive-form Games with Imperfect Recall. *arXiv preprint arXiv:1205.0622*.

Lisý, V.; Davis, T.; and Bowling, M. 2016. Counterfactual Regret Minimization in Sequential Security Games. In *Proceedings of AAAI Conference on Artificial Intelligence*.

Piccione, M., and Rubinstein, A. 1997. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* 20(1):3–24.

Sandholm, T. 2010. The State of Solving Large Incomplete-Information Games, and Application to Poker. *AI Magazine* 13–32.

Waugh, K.; Zinkevich, M.; Johanson, M.; Kan, M.; Schniezel, D.; and Bowling, M. H. 2009. A Practical Use of Imperfect Recall. In *SARA*. Citeseer.

Wichardt, P. C. 2008. Existence of Nash Equilibria in Finite Extensive Form Games with Imperfect Recall: A Counterexample. *Games and Economic Behavior* 63(1):366–369.

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2007. Regret Minimization in Games with Incomplete Information. *NIPS*.